所属类别	**2020 年“华数杯”全国大学生数学建模竞赛**	参赛编号

基于大数据对脱贫帮扶绩效的评价

摘要

今年是全面建成小康社会目标实现之年，也是全面打赢脱贫攻坚战收官之年，任务艰巨又遭遇疫情影响，给打赢脱贫攻坚战增添了困难。为了更好的激励各帮扶单位提高扶贫效率，我国的研究团队分别在五年前与今年对需要帮扶的贫困村进行了贫困调研，本文基于脱贫帮扶绩效评价相关数据，结合实际情况，对于不同问题建立了不同的数学模型进行求解。

针对问题一：以**典型相关分析**法为主要解题思路，将五年前的各项评价指标与今年的各项评价指标作为两组典型变量，建立**典型相关模型**，利用附件中的指标数据得出五个评价指标即居民收入、产业发展、居住环境、文化教育、基础设施的**典型相关系数**分别为0.5342、0.6399、0.7372、0.6361、0.5945。根据典型相关系数判断出居住环境这一指标的相关性较强，其他四个指标的相关性较弱，但总的来说，没有相关性非常强的指标。为了检验该方法的正确性，进行显著性检验，所有的检验结果都是通过，得出结论并该问各指标没有很强烈的相关规律。

针对问题二：以**主成分分析法**为主要解题思路，我们将五个评分数据与总分数据综合考虑，计算六个数据在2020年相较于2015年的增长率，利用主成分分析法，将6个数据的增长率与2020年的6个原始数据作为相关系数矩阵，得到该矩阵的特征值以及相应的特征向量，并根据各个主成分的**贡献率**选取主成分的个数，建立了基于评分数据以及总分数据增长率的**得分模型**，计算出所有贫困村对应的综合得分，根据得分的高低得出不同类型的帮扶单位绩效的高低，给出了脱贫帮扶绩效前十名的帮扶单位编号分别为58、100、20、97、144、145、79、55、58、58。再综合得分排名前列的帮扶单位统计每种单位类型所占比例，得出帮扶单位类型绩效排名为为：类型3>类型1>类型5>类型2。

针对问题三：我们单独选择每个标号所包含单位对应的村庄，将这些村庄的每个指标的增长率求取平均值，然后依次选择0-159各个标号所包含单位对应的村庄，对这些村庄的每个指标的**增长率**求取平均值，建立基于增长率年平均值的**排序模型**，将各个标号所对应该指标的增长率平均值进行排序，得出了各单项评价指标帮扶业绩明显的帮扶单位。

针对问题四：我们选择 2015 年的六个指标数据以及帮扶单位的编号为输入，2020 年的总分数据为输出，利用**神经网络算法**，建立**预测模型**，预测出该 10 个村庄 2020 年的总分数据，将其与总分数据排名前 10000 名的村庄相比较，判断出编号为 25149 的村庄能评上“脱贫先进村庄”称号。根据一级称号与二级称号的比例判断出编号为 25149 的村庄为 “脱贫先进一级村庄”。

针对问题五：我们向国家扶贫办写了一封建议信，阐述了我们对脱贫攻坚的认识并给出了相关的观点和建议。

关键词：**相关系数　主成分分析法　得分模型　排序模型　神经网络算法　预测模型**

一、 问题重述

1.1 问题背景

消除贫困、改善民生、逐步实现共同富裕，是社会主义的本质要求，是我们党的重要使命。党的十八大以来，国家把扶贫开发工作纳入“四个全面”战略布局，作为实现第一个百年奋斗目标的重点工作，摆在更加突出的位置。

党的十八大以来，以习近平同志为核心的党中央采取一系列超常规举措，举全党全社会之力打赢脱贫攻坚战，使得我国扶贫工作不断取得新的进展，其中所积累的一系列先进经验值得深刻总结，包括“以人民为中心”的理念贯穿始终、坚持制度变革推动减贫进程、坚持以经济发展推动贫困治理等等。上述基本经验对于探寻我国扶贫工作的基本规律、开展 2020 年后治理相对贫困工作具有重要意义[1]。

2020 年是精准扶贫的收官之年,如何防止贫困反弹并实现扶贫效果长久化、制度化,成为政府及社会各界必须面对和考虑的战略性问题。研究指出,“后 2020”时期,绝对贫困即将成为历史,相对贫困将会显现,为应对相对贫困,需要调整扶贫标准,克服扶贫过程中暴露出来的一些痼疾顽症[2]。因此，为了了解我国的脱贫情况，我们有必要利用相关数学算法，结合大数据背景，评价当前脱贫帮扶的绩效如何，从而提出针对性的应对措施。

1.1 问题提出

为了更好地评价脱贫帮扶绩效，我们建立数学模型解决下列问题：

（1）一般人的理解是，五年前的评分与2020 年对应的各项评分有着直接的关联，如五年前的居民收入不错，现在的居民收入也会是不错的。你认为本问题有这个规律吗？请分析各个评价指标的对应关系。

（2）160个帮扶单位帮扶着基础不同的村庄，帮扶单位帮扶工作的态度、目标、投入、帮扶干部素质等显然是有差异的。仅仅用2020年各村庄评分高低显然是无法真正有效的体现一个帮扶单位在脱贫攻坚提升方面所做出的努力。请你运用附件的数据，阐明什么类型的帮扶单位，哪些帮扶单位在脱贫帮扶上面有较高的绩效？请给不同类型的帮扶单位绩效排序，给出脱贫帮扶绩效前十名的帮扶单位编号。

（3）每个帮扶单位在扶贫上有不同的工作特色，如有些单位在提高居民收入上效果很好，而有些帮扶单位可能在改善基础设施上帮助的效果不错。请问，哪些帮扶单位分别在居民收入、产业发展、居住环境、文化教育、基础设施等评价指标上帮扶业绩明显？请列出各单项评价指标前五名的帮扶单位编号。

（4）全国计划给予10000个村庄“脱贫先进村庄”称号。请问，哪些因素对获得这个荣誉称号有着非常重要的影响？数据表中最后有10个村庄的2020年的评价分数被删除，请你判断他们能评上“脱贫先进村庄”称号吗？如果称号分为一级和二级（一、二级称号比例为1:3），这10个村庄中谁能评上“脱贫先进一级村庄”称号？

（5）依据你的研究成果，向国家扶贫办写一封500字左右的信，阐述你的观点和建议。

二、问题分析

问题一：该问题要求我们判断各个评价指标五年前的评分与 2020 年对应的评分是否有着直接的关联，并分析各个评价指标的对应关系。这是一个对指标进行相关性判断的问题，我们可以以典型相关分析法为主要思路，将五年前的各项评价指标与今年的各

项评价指标作为两组典型变量，利用附件中的指标数据得出五个评价指标典型相关系数，取某一数值为典型相关系数的临界值，若某一指标的相关系数大于该值，则该指标五年前的评分与 2020 年对应的各项评分有着直接的关联，否则认为该指标五年前的评分与 2020 年对应的各项评分没有直接的关联。为了检验该方法的正确性，可进行显著性检验，若所有的检验结果都是通过，则证明使用典型相关分析法的正确性。为分析各个评价指标的对应关系，可建立典型相关模型，基于典型相关系数的区间范围，将各个评价指标的对应关系进行分类，该相关系数越大，关系越密切，从而判断出各个评价指标的对应关系。

问题二：该问题要求我们运用附件的数据，阐明什么类型的帮扶单位，哪些帮扶单位在脱贫帮扶上面有较高的绩效，并给不同类型的帮扶单位绩效排序，给出脱贫帮扶绩效前十名的帮扶单位编号。我们选择将五个评分数据与总分数据综合考虑，计算六个数据在 2020 年相较于 2015 年的增长率，利用主成分分析法，将 6 个数据的增长率与 2020 年的 6 个原始数据作为相关系数矩阵，并根据各个主成分的贡献率选取主成分的个数，计算出所有贫困村对应的综合得分，根据得分的高低便可得出不同类型的帮扶单位绩效的高低，对于得分高的帮扶单位即为在脱贫帮扶上面有较高绩效的帮扶单位。再综合得分排名前列的帮扶单位统计每种单位类型所占比例，数量比例高的即为在脱贫帮扶上面有较高绩效的帮扶单位类型。

问题三：问题三要求我们列出各单项评价指标前五名的帮扶单位编号，由于被帮扶的村庄划分为160个集合，每个集合指定帮扶单位（标记为0-159）进行帮扶，即每个标号对应一个帮扶单位，该帮扶单位分别帮扶其对应的贫困村庄。我们可以单独选择每个标号所包含单位对应的村庄，将这些村庄的每个指标的增长率求取平均值，然后依次选择0-159各个标号所包含单位对应的村庄，对这些村庄的每个指标的增长率求取平均值。这样一来，就某一单项评价指标而言，将各个标号所对应该指标的增长率平均值进行排序，从而得出针对某一单项评价指标帮扶业绩明显的帮扶单位。

问题四：问题四首先要求我们给出影响“脱贫先进村庄”这一称号的因素，为了判断最后10个村庄能否评上该称号，我们可以选择2015年的六个指标数据以及帮扶单位的编号为输入，2020年的总分数据为输出，利用神经网络算法，建立预测模型，通过预测出的该10个村庄2020年的总分数据，与总分数据排名前10000名的村庄相比较，从而判断他们是否能评上该称号。由于一级称号与二级称号的比例为1：3，选择总分数据前2500名的村庄获得一级称号，总分数据后7500名的村庄获得二级称号。那么这10个村庄中总分数据位于前2500名则能评上“脱贫先进一级村庄”的称号。

问题五：问题五要求我们根据给国家扶贫办写一封信，阐述我们团队对于当前我国脱贫攻坚的认识，并给出一些观点和建议。结合前四个问题的研究结果，并结合我国的实际情况，向国家扶贫办给出一些对于扶贫开发工作的观点和建议。

三、基本假设

1、假设对于脱贫帮扶绩效评价的标杆，模型所确立的指标，能够准确地代表脱贫帮扶绩效的高低；

2、假设题目附件所给的科研团队调研数据真实有效；

3、假设科研团队在进行贫困调研时，对各个村庄的评价公平公正；

4、假设不考虑贫困村的人口流动等因素的影响，对于脱贫帮扶绩效的评价，除去本文提到的指标因素，对其他因素所造成的影响忽略不计；

5、假设在不考虑其他因素的情况下，某村庄的总分数据能够反映该村庄目前的贫困情

况；
6、在打分过程中，各因子的贡献率能够代表其在实际生活所占权重。

四、符号说明

符号	说明
U_k ， V_k	某一指标的第 k 对典型相关变量
ρ_k	第 k 对典型相关变量的典型相关系数
e_k, f_k	正交单位特征向量
a	居民收入的增长率
b	产业发展的增长率
c	居住环境的增长率
d	文化教育的增长率
e	基础设施的增长率
f	总分的增长率
b_j	信息贡献率
α_p	累积贡献率
Z	综合得分

五、数据的预处理

5.1数据的筛选和删除

通过查询网站数据并结合附件所给数据分析，可以将一些无关项删除，保留有用项。如通信网络、生产条件、社会保障等指标，由于这些指标数据对研究项目影响很小或者没有，所以本文仅选择附件中的数据将其保留。

5.2 数据的标准化

理论上，标准化适用于服从正态分布的数据，目前很多工程都依赖大数据，所以在样本足够多的情况下，往往直接使用标准化对数据进行无量纲化预处理，在深度学习中，将数据标准化能够保证有更好的收敛。如果不进行数据标准化，有些特征将会对损失函数影响很大，使得其他值比较小的特征重要性降低。

为了便于比较和研究，本文已将所有数据都进行了标准化处理，故不再作处理。

六、模型的建立与求解

6.1问题一模型的建立与求解

6.1.1问题一模型的建立

问题一要求我们分析各个评价指标的对应关系，我们将五年前的各项评价指标与今年的各项评价指标作为两组典型变量，利用附件中的指标数据得出五个评价指标典型相关系数，建立典型相关模型，取0.5为典型相关系数的临界值，若某一指标的相关系数大于该值，则该指标五年前的评分与2020年对应的各项评分有着直接的关联，否则认为该指标五年前的评分与2020年对应的各项评分没有直接的关联，从而得出各个评价指标的对应关系。

•模型的准备

通过查阅文献可知[3]，对于两组随机变量$(x_1,x_2,...,x_p)$和$(y_1,y_2,...,y_p)$，考虑$(x_1,x_2,...,x_p)$的一个线性组合U及$(y_1,y_2,...,y_p)$的一个线性组合V,希望找到U和V之间的最大可能的相关系数，来充分反映两组变量之间的关系。这样就可以把研究两组随机变量间相关关系的问题转化为研究两个随机变量之间的相关关系。如果说一对变量（U,V）还不能完全刻画两组变量间的相关关系，可以继续找第二对变量，但是要保证第二对的变量与第一对的变量不相关，直至找不到相关变量时为止。

（1）确定总体典型相关变量

设有两组随机向量$x=(x_1,x_2,...,x_p)^T$，$y=(y_1,y_2,...,y_q)^T(p\le q)$，将两组合并成一组向量$(x^T,y^T)=(x_1,x_2,...,x_p,y_1,y_2,...,y_q)^T$，其协方差矩阵为：

$$\sum=\begin{bmatrix}\sum_{11}\sum_{12}\\ \sum_{21}\sum_{22}\end{bmatrix} \qquad (6\text{-}1\text{-}1)$$

其中$\sum_{11}=\mathrm{cov}(x)$，$\sum_{22}=\mathrm{cov}(y)$，$\sum_{12}=\sum_{21}{}^T=\mathrm{cov}(x,y)$。

根据典型相关思想，问题是要寻找$x=(x_1,x_2,...,x_p)^T, y=(y_1,y_2,...,y_q)^T(p\le q)$的线性组合：

$$U_1=a_1^Tx=a_{11}x_1+a_{12}x_2+...+a_{1p}x_p$$
$$V_1=b_1^Ty=b_{11}y_1+b_{12}y_2+...+b_{1q}y_q$$

要使得U_1,V_1的相关系数$\rho(U_1,V_1)$达到最大，这里

$$a_1^T=(a_{11},a_{12},...,a_{1p})$$
$$b_1^T=(b_{11},\ b_{12},...,b_{1q})$$

由（6-1-1）式，$\mathrm{var}(U_1)=a_1^T\sum_{11}a_1$，$\mathrm{var}(V_1)=b_1^T\sum_{22}b_1$，$\mathrm{cov}(U_1,V_1)=a_1^T\sum_{12}b_1$，所以$U_1,V_1$的相关系数为：

$$\rho(U_1,V_1)=\frac{a_1^T\sum_{12}b_1}{\sqrt{a_1^T\sum_{11}a_1}\sqrt{b_1^T\sum_{22}b_1}} \qquad (6\text{-}1\text{-}2)$$

又由于相关系数与量纲无关，因此可设定约束条件：

$$a_1^T\sum_{11}a_1=b_1^T\sum_{22}b_1=1 \qquad (6\text{-}1\text{-}3)$$

满足约束条件（6-1-3）的相关系数 $\rho(U_1,V_1)$ 的最大值称为第一典型相关系数，U_1,V_1 称为典型相关变量。

（2）计算典型相关系数

将矩阵 $[x^T,Y^T]^T$ 的协方差矩阵或相关系数矩阵表示如下：

$$\sum=\begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix},\quad R=\begin{bmatrix}R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix}$$

令 $A=(\sum_{11})^{-1/2}\sum_{12}(\sum_{22})^{-1}\sum_{21}(\sum_{11})^{-1/2}$，$B=(\sum_{22})^{-1/2}\sum_{21}(\sum_{22})^{-1}\sum_{12}(\sum_{22})^{-1/2}$，求 A,B 的特征值 ρ_1^2，ρ_2^2，$\ldots$，ρ_p^2 以及对应的正交单位特征向量 e_k，$f_k(k=1,2,\ldots,p)$。

则 x,y 的第 k 对典型相关变量为：

$$\begin{cases}U_k=a_k^Tx=e_k^T\sum_{11}^{-0.5}x\\ V_k=b_k^Ty=f_k^T\sum_{22}^{-0.5}y\end{cases},k=1,2,\ldots,p$$

其中 $\sum_{11}^{-0.5}$，$\sum_{22}^{-0.5}$ 分别为 $\sum_{11}$，$\sum_{22}$ 的平方根矩阵的逆矩阵。

x,y 的第 k 对典型相关变量的相关系数为：

$$\rho_k=a_k^T\sum_{12}b_k(k=1,2,\ldots,p)$$

• 模型的建立

由典型相关分析可知，将2015年的居民收入(SR)、产业发展(CY)、居住环境(HJ)、文化教育(WJ)、基础设施(SS)五个评价指标分别与2020年的居民收入、产业发展、居住环境、文化教育、基础设施作为五组典型相关变量，建立典型相关模型，画出模型结构如下：

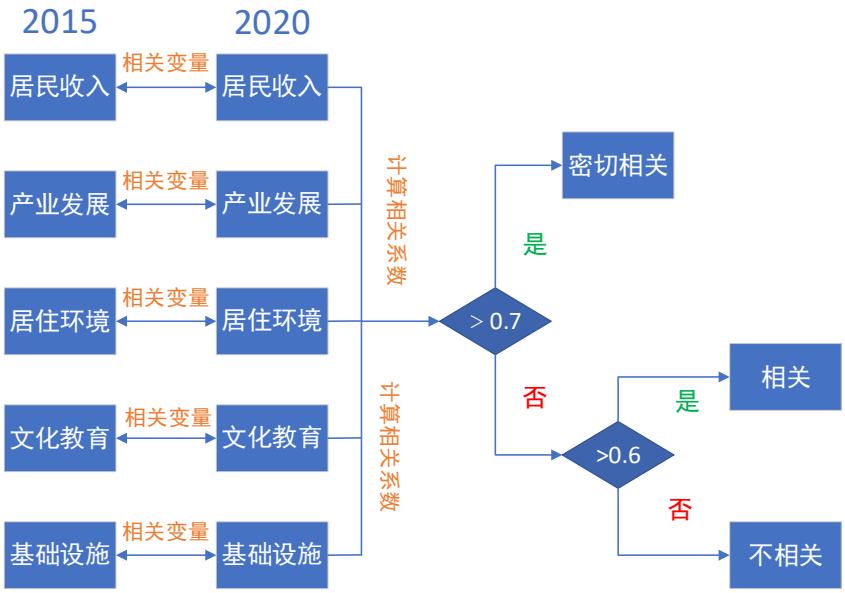

图1：典型相关模型结构图

由上图可知各个评价指标最终的对应关系是由典型相关系数来判断的，相关系数越

大，说明该指标在五年前与今年的关联度越高，该系数大小的确定需要根据题目中所给出的数据进行求解，求解过程及结果在下面模型求解中给出。至此，我们建立出典型相关模型如下：

$$\begin{cases} U_k = a_k^T x = e_k^T \sum_{11}^{-0.5} x \\ V_k = b_k^T y = f_k^T \sum_{22}^{-0.5} y \end{cases} (k = 1,2,...,32165) \qquad (6\text{-}1\text{-}4)$$

$$\rho_k = a_k^T \sum_{12} b_k (k = 1,2,...,32165) \qquad (6\text{-}1\text{-}5)$$

其中U_k，V_k为某一指标的第k对典型相关变量，ρ_k第k对典型相关变量的典型相关系数。

相关系数的意义如下表所示：

表1：相关系数的意义

相关系数绝对值	大体划分	细分
0.9-1.0	相关	相关性非常强
0.7-0.9	相关	相关性比较强
0.5-0.7	相关	相关性有点弱
<0.5	不相关	相关性非常弱

6.1.2 问题一模型的求解

该问题的整体求解思路为计算出各个评价指标的典型相关系数，再对结果进行显著性检验，检验该典型相关分析法是否正确。那么模型的求解可大致分为以下 3 步：

Step1：求解典型相关系数

根据题目附件所给的五个评价指标的相关数据，利用协方差矩阵和相关系数矩阵求得各指标的典型相关系数如下表：

表 2：各指标的典型相关系数

居民收入（SR）	产业发展（CY）	居住环境（HJ）	文化教育（WJ）	基础设施（SS）	总分
0.5342	0.6399	0.7372	0.6361	0.5945	0.7845

Step2：显著性检验

在本模型中我们使用的检验方法是 P 值检验与统计量检验，P 值即为(零假设的)拒绝域的面积或概率。用 P 值做判断与检验统计量做判断得出的结论是完全一样的，在检验中有两种方法，第一种，检验统计量是否落在拒绝域中，第二种，P 值是否小于显著性水平。

其实两种判断是等价的，拒绝域是根据显著性水平计算出来的，然后看检验统计量与拒绝域比较；而 P 值是根据检验统计量算出来的，然后看 P 值与显著性水平比较。在

分布图上，P 值所在位置就是检验统计量所在位置。

表 3：显著性检验表

wilks	df1	df2	F	pF	Chisq	pChisq	dfe	p
0.7146	1	32153	1.2842e+04	0	1.0805e+04	0	1	0
0.5905	1	32153	2.2301e+04	0	1.6939e+04	0	1	0
0.4566	1	32153	3.8267e+04	0	2.5206e+04	0	1	0
0.5954	1	32153	2.1853e+04	0	1.6674e+04	0	1	0
0.6466	1	32153	1.7574e+04	0	1.4020e+04	0	1	0
0.3846	1	32153	5.1442e+04	0	3.0721e+04	0	1	0

其中，wilks 为似然比统计量，df1 为 F 统计量的分子自由度（卡方统计的自由度），df2 为 F 统计量的分母自由度，F 为近似 F 统计量，pF 为 F 的右尾显著性水平，chisq 为近似卡方统计，pchisq 为 chisq 的右尾显著性水平。

Step3：得出各个评价指标的对应关系

由表 2 可知，第三个指标（居住环境）的典型相关系数位于区间[0.7,0.9]，表明五年前的评分与 2020 年对应的各项评分相关，且相关性较强；第一个指标（居民收入）、第二个指标（产业发展）、第四个指标（文化教育）与第五个指标（基础设施）的典型相关系数位于区间[0.5,0.7]，表明五年前的评分与 2020 年对应的各项评分相关，且相关性较弱。

由表 3 可知，五对典型变量均通过了显著性检验，表明相应典型变量之间相关关系显著。

因此，我们得出结论：对于本问中所提到的五年前的评分与2020 年对应的各项评分有着直接的关联这一规律，居住环境具有这种规律，关联度较高；居民收入、产业发展、文化教育与基础设施也具有这种规律，但关联度较低。总的来说，没有相关性非常强的指标。

6.2 问题二模型的建立与求解

6.2.1 问题二模型的建立

问题二要求我们运用附件的数据，阐明什么类型的帮扶单位，哪些帮扶单位在脱贫帮扶上面有较高的绩效，并给不同类型的帮扶单位绩效排序，给出脱贫帮扶绩效前十名的帮扶单位编号。我们将五个评分数据与总分数据综合考虑，计算六个数据在2020年相较于2015年的增长率，利用主成分分析法，将6个数据的增长率与2020年的6个原始数据作为相关系数矩阵，计算出该矩阵的特征值，并根据各个主成分的贡献率选取主成分的个数，建立基于评分数据及总分数据增长率的得分模型。根据得分的高低便可得出不同类型的帮扶单位绩效的高低，对于得分高的帮扶单位即为在脱贫帮扶上面有较高绩效的帮扶单位。

•模型的准备

主成分分析法是最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中，并期望在所投影的维度上数据的信息量最大（方差最大），以此使用较少的数据维度，同时保留住较多的原数据点的特性。

假设有 M 个样本 $\{x_1, x_2, x_3, ..., x_m\}$，每个样本有 N 维特征 $x^i = (x_1^i, x_2^i, x_3^i, ..., x_N^i)^T$，每一个特征 x_j 都有各自的特征值。

（1）对特征进行中心化

求每一个特征的平均值，然后对于所有的样本，每一个特征都减去自身的均值。特征 x_j 的平均值为：

$$\overline{x_j} = \frac{1}{M}\sum_{i=1}^{M} x_j^i \qquad (6\text{-}2\text{-}1)$$

经过去均值处理之后，原始特征的值就变成了新的值。在此基础上，进行后面的操作。

（2）求协方差矩阵 C

$$C = \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \mathrm{cov}(x_1, x_2) \\ \mathrm{cov}(x_2, x_1) & \mathrm{cov}(x_2, x_2) \end{bmatrix}$$

上述矩阵中，对角线上分别是特征 x_1 和 x_2 的方差，非对角线上是协方差。协方差大于 0 表示 x_1 和 x_2 若有一个增，另一个也增；小于 0 表示一个增，一个减；协方差为 0 时，两者独立。协方差绝对值越大，两者对彼此的影响越大，反之越小。其中，$\mathrm{cov}(x_1, x_2)$ 的求解公式如下：

$$\mathrm{cov}(x_1, x_2) = \frac{\sum_{i=1}^{M}(x_1^i - \overline{x_1})(x_1^i - \overline{x_1})}{M - 1} \qquad (6\text{-}2\text{-}2)$$

其他可类似进行求解，根据上面的协方差计算公式我们就得到了这 M 个样本在这 N 维特征下的协方差矩阵 C。

（3）求协方差矩阵 C 的特征值和特征向量

利用矩阵的知识，求协方差矩阵 C 的特征值 λ 和相对应的特征向量 u：

$$Cu = \lambda u$$

特征值 λ 会有 N 个，每一个 λ_i 对应一个特征向量 u_i，将特征值 λ 按照从大到小的顺序排序，选择最大的前 k 个，并将其相对应的 k 个特征向量拿出来，我们会得到一组 $\{(\lambda_1, u_1), (\lambda_2, u_2), ..., (\lambda_k, u_k)\}$。

（4）得到降维后的新 k 维特征

这个选取最大的前 k 个特征值和相对应的特征向量，并进行投影的过程，就是降维的过程。对于每一个样本，原来的特征是 $(x_1^i, x_2^i, x_3^i, ..., x_N^i)^T$，投影之后的新特征是 $(y_1^i, y_2^i, y_3^i, ..., y_k^i)^T$，新特征的计算公式如下：

$$\begin{bmatrix} y_1^i \\ y_2^i \\ . \\ . \\ . \\ y_k^i \end{bmatrix} = \begin{bmatrix} u_1^T \cdot (x_1^i, x_2^i, ..., x_n^i)^T \\ u_2^T \cdot (x_1^i, x_2^i, ..., x_n^i)^T \\ ..... \\ ..... \\ ..... \\ u_k^T \cdot (x_1^i, x_2^i, ..., x_n^i)^T \end{bmatrix}$$

•模型的建立

我们将五个评分数据与总分数据综合考虑，计算六个数据在 2020 年相较于 2015 年的增长率，利用主成分分析[4]法，将 6 个数据的增长率与 2020 年的 6 个原始数据作为相关系数矩阵，计算出该矩阵的特征值，并根据各个主成分的贡献率选取主成分的个数，建立基于评分数据及总分数据增长率的得分模型。根据各个帮扶单位得分的高低便可得出不同类型的帮扶单位绩效的高低，对于得分高的帮扶单位即为在脱贫帮扶上面有较高绩效的帮扶单位。

下面给出该得分模型的流程图：

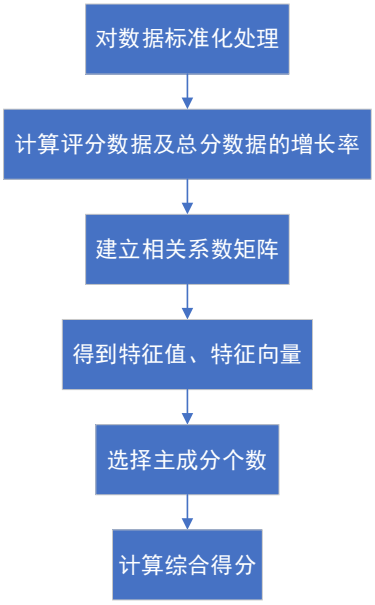

图 2：得分模型流程图

我们对所有贫困村基于其评分数据与总分数据在 2020 年相较于 2015 年的增长率计算出各个村庄的综合得分，得分高的贫困村所对应的帮扶单位在脱贫帮扶上面有较高的绩效。将所有贫困村的得分进行排序，则得到了帮扶单位的绩效排序。我们选取综合得分位于前一百名的贫困村所对应的帮扶单位，认为他们在脱贫帮扶上面有较高的绩效；选取综合得分位于某一范围内的贫困村所对应的帮扶单位，计算不同类型帮扶单位在该范围内的数量比例，认为比例较高的帮扶单位类型在脱贫帮扶上面有较高的绩效。

至此，我们建立基于评分数据及总分数据增长率的得分模型如下：

（1）求取评分数据及总分数据的增长率

令居民收入、产业发展、居住环境、文化教育、基础设施评分以及总分的增长率分别为a、b、c、d、e、f，表示如下：

$$\begin{cases} a=\dfrac{SR(2020)-SR(2015)}{SR(2015)} \\ b=\dfrac{CY(2020)-CY(2015)}{CY(2015)} \\ c=\dfrac{HJ(2020)-HJ(2015)}{HJ(2015)} \\ d=\dfrac{WJ(2020)-WJ(2015)}{WJ(2015)} \\ e=\dfrac{SS(2020)-SS(2015)}{SS(2015)} \\ f=\dfrac{\text{总分}(2020)-\text{总分}(2015)}{\text{总分}(2015)} \end{cases} \qquad (6\text{-}2\text{-}3)$$

（2）建立相关系数矩阵 R

根据我们选取的数据分析可知，相关系数矩阵 $R=(r_{ij})_{32165\times12}$。

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdot & \cdot & r_{(1)(12)} \\ r_{21} & r_{22} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{(32165)(1)} & \cdot & \cdot & \cdot & r_{(32165)(12)} \end{bmatrix}$$

其中，当 $j\le6$ 时，r_{ij} 为第 i 个贫困村第 j 个数据的增长率；当 $j>6$ 时，r_{ij} 为第 i 个贫困村 2020 年第（j-6）个数据的原始值。

（3）计算特征值和特征向量

计算相关系数矩阵 R 的特征值 $\lambda_j=(\lambda_1,\lambda_2,...,\lambda_{12})$ 以及对应的标准化特征向量 $u_j=(u_1,u_2,...,u_{12})^T$，由特征向量组成 12 个新的指标变量：

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_{12} \end{bmatrix}=\begin{bmatrix} u_1\cdot(x_1,x_2,...,x_{12}) \\ u_2\cdot(x_1,x_2,...,x_{12}) \\ ..... \\ ..... \\ ..... \\ u_{12}\cdot(x_1,x_2,...,x_{12}) \end{bmatrix}$$

（4）选择主成分个数

通过计算特征值 $\lambda_j(j=1,2,...,12)$ 的信息贡献率 b_j 与累积贡献率 α_p，两者的计算方法如下：

$$\begin{cases} b_j = \dfrac{\lambda_j}{\sum_{k=1}^{12}\lambda_k}, j = 1,2,...,12 \\ \alpha_p = \dfrac{\sum_{k=1}^{p}\lambda_k}{\sum_{kk=1}^{12}\lambda_k} \end{cases} \qquad (6\text{-}2\text{-}4)$$

当α_p接近于 1 时，则选择前 p 个指标变量 $y_1, y_2, ..., y_p$ 作为 p 个主成分，代替原来的指标变量，从而可对 p 个主成分进行综合分析。

（5）计算综合得分 Z

最后得出的综合得分值作为我们评判帮扶单位绩效高低的标准，得分越高，代表该贫困村脱贫情况越好，即该贫困村所对应的帮扶单位绩效越高。

$$Z = \sum_{j=1}^{p} b_j y_j \qquad (6\text{-}2\text{-}5)$$

其中，b_j 为第 j 个主成分的信息贡献率，根据综合得分值就可进行相应排序与评价。

6.2.2 问题二模型的求解

我们利用计算机软件对32165个贫困村的评分数据以及总分数据算出各自指标的增长率，通过构建相关系数矩阵得到特征值如下表所示：

表 4：矩阵特征值

	1	2	3	4	5	6
特征值	4.7110	1.0283	1.0216	0.9960	0.9835	0.9697
	7	8	9	10	11	12
特征值	0.9666	0.5128	0.3168	0.2545	0.2308	0.0079

由于通过计算机软件的特征向量数据过于冗杂，故这里展示部分结果，完整结果详见附录。

表 5；部分特征向量值

1	2	3	4
0.0372	0.6433	0.2117	0.4825
0.0670	0.2869	0.1815	0.0480
0.0410	0.0459	0.6817	0.2636
0.0476	0.2664	0.2631	0.8044

由式(6-2-4)分别计算出 12 个特征值的贡献率如下表所示：

表 6：特征值的贡献率

	1	2	3	4	5	6
贡献率	39.2584	8.5692	8.5139	8.3005	8.1964	8.0813
	7	8	9	10	11	12

贡献率	8.0550	4.2735	2.6404	2.1214	1.9238	0.0662

最终，我们计算出所有贫困村基于评分数据以及总分数据的的综合得分，这里展示得分前十五名的贫困村庄排名，完整排名详见附件“sj2.xlsx”。

表 7：部分贫困村综合得分排名

村庄编号	帮扶单位 (0-159)	帮扶单位类型 (0-5)	综合得分
24032	58	1	75.40644823
28745	100	5	66.5621569
15344	20	2	60.97766683
28103	97	3	56.14412667
50096	144	3	53.93653142
50399	145	3	47.74866269
34386	79	3	45.76284147
23490	55	2	45.19598216
24041	58	1	44.80786136
23945	58	1	43.75222852
29998	104	3	39.99029972
51595	148	3	38.48604738
28100	97	3	37.41162991
25823	91	3	34.63453939
25965	91	3	33.70653591

由上表可知，脱贫帮扶绩效前十名的帮扶单位编号分别为58、100、20、97、144、145、79、55、58、58。

对于阐明哪些帮扶单位在脱贫帮扶上面有较高的绩效这一问题，我们根据附件“sj2.xlsx”选取综合得分位于前一百名的贫困村所对应的帮扶单位，认为他们在脱贫帮扶上面有较高的绩效。

对于阐明什么类型的帮扶单位在脱贫帮扶上面有较高的绩效这一问题，我们选取综合得分位于某一范围内的贫困村所对应的帮扶单位，计算不同类型帮扶单位在该范围内的数量比例，认为比例较高的帮扶单位类型在脱贫帮扶上面有较高的绩效。

我们选取了两种情况进行评价分析：

① 对得分排名前15的贫困村分析

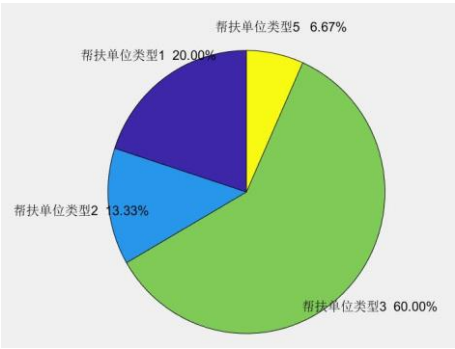

图3：前15名不同帮扶单位类型数量比例

由图3可知，帮扶单位类型3所占比例最大，按数量比例从大到小的顺序排列，得分位于前15名的贫困村所对应的帮扶单位类型数量为：类型3>类型1>类型5>类型2。

② 对得分排名前50的贫困村分析

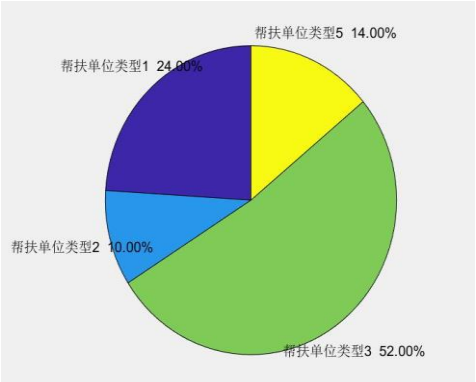

图4：前50名不同帮扶单位类型数量比例

由图 4 可知，依然是帮扶单位类型 3 所占比例最大，按数量比例从大到小的顺序排列，得分位于前 50 名的贫困村所对应的帮扶单位类型数量为：类型 3>类型 1>类型 5>类型 2。

综合以上结论，我们认为帮扶单位类型 3 在脱贫帮扶上面有较高的绩效。

6.3 问题三模型的建立与求解

6.3.1 问题三模型的建立

问题三要求我们列出各单项评价指标前五名的帮扶单位编号，由于被帮扶的村庄划分为160个集合，每个集合指定帮扶单位（标记为0-159）进行帮扶，即每个标号对应一个帮扶单位，该帮扶单位分别帮扶其对应的贫困村庄。通过附件中的数据，单独选择每个标号所包含单位对应的村庄，将这些村庄的每个指标的增长率求取平均值，同理，依次选择0-159各个标号所包含单位对应的村庄，对这些村庄的每个指标的增长率求取平均值。这样一来，就某一单项评价指标而言，将各个标号所对应该指标的增长率平均值进行排序，建立基于单项指标增长率平均值的排序模型，从而得出针对某一单项评价指标帮扶业绩明显的帮扶单位。

该排序模型流程图如下：

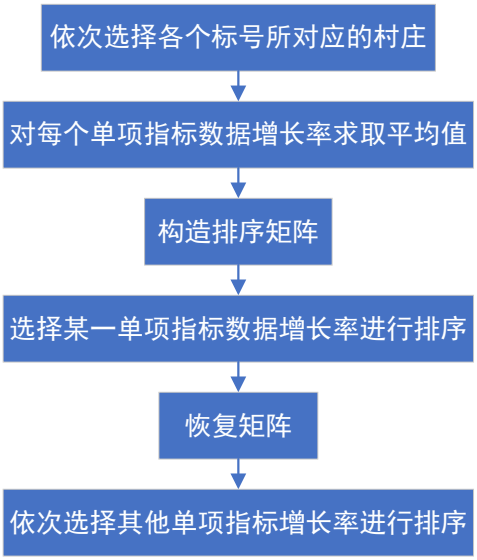

图5：排序模型流程图

至此，我们建立出建立基于单项指标增长率平均值的排序模型如下：

（1）构造排序矩阵

设标号0-159所对应村庄的居民收入(SR)增长率平均值分别为$\alpha_{SR0},\alpha_{SR1},...,\alpha_{SR159}$，产业发展(CY)增长率平均值分别为$\alpha_{CY0},\alpha_{CY1},...,\alpha_{CY159}$，居住环境(HJ)增长率平均值分别为$\alpha_{HJ0},\alpha_{HJ1},...,\alpha_{HJ159}$、文化教育(WJ)增长率平均值分别为$\alpha_{WJ0},\alpha_{WJ1},...,\alpha_{WJ159}$、基础设施(SS)增长率平均值分别为$\alpha_{SS0},\alpha_{SS1},...,\alpha_{SS159}$。

结合式（6-2-3）增长率的计算公式，将标号与各个指标的增长率平均值构造成一个160x6的排序矩阵：

$$\begin{bmatrix} 0 & \alpha_{SR0} & \alpha_{CY0} & ... & \alpha_{SS0} \\ 1 & \alpha_{SR1} & \alpha_{CY1} & ... & \alpha_{SS1} \\ 2 & \alpha_{SR2} & ... & ... & ... \\ ... & ... & ... & \alpha_{WJ158} & \alpha_{SS158} \\ 159 & ... & ... & \alpha_{WJ159} & \alpha_{SS159} \end{bmatrix}$$

（2）选取单项指标进行排序

对每个单项评价指标而言，从排序矩阵中单独选取该指标所对应得那一列，利用计算机软件对该列的增长率平均值α进行降序排序，根据排序结果即可得出各单项评价指标帮扶业绩明显的帮扶单位。

6.3.2 问题三模型的求解

该问题在前两问的基础上，进行了更加细致的评价，选择了单项指标对帮扶单位的绩效进行评价，对于该问题的求解过程大致分为以下 2 步：

Step1：得出排序矩阵

由于数据过于繁多，部分矩阵如下表所示，完整矩阵详见附件：“jieguo3.xlsx”。

表8：部分排序矩阵

0	**-0.16477**	**0.016326**	**0.329099**	**0.055403**	**0.234771**	**0.023215**
1	0.017693	-0.02951	0.15246	-0.13817	-0.10018	-0.11234
2	-0.23659	-0.27092	-0.21418	-0.26374	-0.27721	-0.31582
3	-1.1616	0.666466	-0.53598	-0.60321	-0.47542	-0.4829
4	-0.26677	-0.92658	-1.2465	-1.00183	-1.30357	-0.94967
5	0.121446	0.483491	0.568274	-0.94436	-0.89895	-0.82682
6	-0.16144	-1.16191	-4.43454	-1.10761	-5.36079	-2.81736
7	-0.74315	1.156745	0.200018	-1.24971	-0.78407	-0.222
…	…	…	…	…	…	…
159	-0.77246	-1.1631	0.679493	0.02203	-0.31819	-0.0055

Step2：选择单项指标进行排序

我们用2020年的指标值减去2015年的指标值，得到每一个指标的差，差值越大说明该类帮扶单位在对应的指标上绩效高，得到差值图示如下：

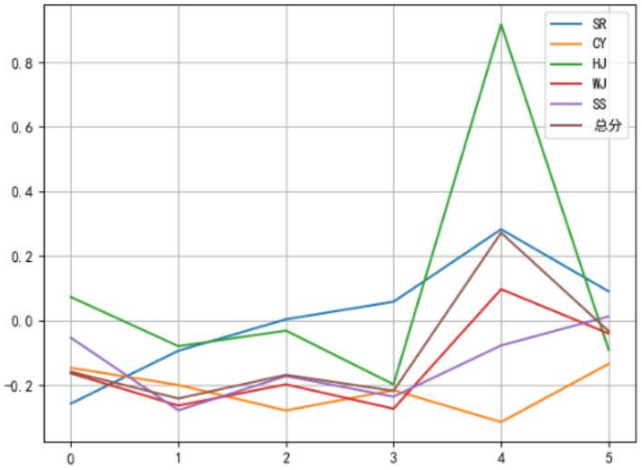

图 6：指标差值

其次我们对每一指标进行具体排序分析，依次列出每项指标的帮扶单位排序结果如下：

① 居民收入

通过计算机软件对居民收入这一指标的增长率平均值进行排序，部分排序结果如下：

表9：居民收入的部分单位排序

20	3.812692
106	3.086956
19	3.034033
55	2.900876
16	2.619129
41	2.384573
29	2.012194
101	1.859347

由上表可知，居民收入这一单项评价指标前五名的帮扶单位编号为20、106、19、55、16。

② 产业发展

通过计算机软件对居民收入这一指标的增长率平均值进行排序，部分排序结果如下：

表10：产业发展的部分单位排序

80	3.534779
157	2.521601
118	2.085688
85	1.904304
102	1.737978
19	1.603025
27	1.370579
45	1.348389

由上表可知，产业发展这一单项评价指标前五名的帮扶单位编号为80、157、118、85、102。

③ 居住环境

通过计算机软件对居民环境这一指标的增长率平均值进行排序，部分排序结果如下：

表11：居住环境的部分单位排序

18	1.368631
154	1.298602
73	1.055104
59	0.704038
65	0.703499
159	0.679493
82	0.602389
5	0.568274

由上表可知，居住环境这一单项评价指标前五名的帮扶单位编号为18、154、73、59、65。

④ 文化教育

通过计算机软件对文化教育这一指标的增长率平均值进行排序，部分排序结果如下：

表12：文化教育的部分单位排序

8	3.843607
72	2.590973
155	1.207743

76	0.876415
19	0.830434
73	0.78587
17	0.778511
55	0.631279

由上表可知，文化教育这一单项评价指标前五名的帮扶单位编号为8、72、155、76、19。

⑤ 基础设施

通过计算机软件对基础设施这一指标的增长率平均值进行排序，部分排序结果如下：

表13：基础设施的部分单位排序

75	1.017953
72	0.745033
94	0.670114
89	0.345746
107	0.289836
146	0.286788
66	0.282462
45	0.24861

由上表可知，基础设施这一单项评价指标前五名的帮扶单位编号为75、72、94、89、107。

至此，我们列出了各单项评价指标前五名的帮扶单位编号，对结果进行进一步的分析，这160个帮扶单位按照单位属性（如国企还是民营企业等）标记为0-5这6个类型，我们对这6个类型的帮扶单位属性选取排序位于前列的进行数量统计，得到各单项评价指标绩效较好的单位类型，对应的饼状图如下图所示：

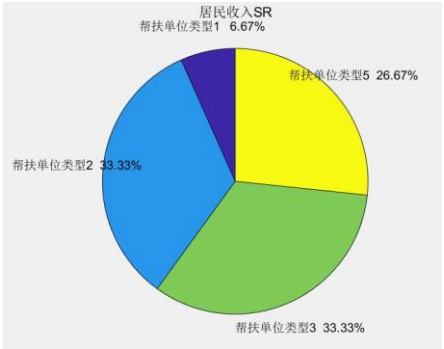

图7：居民收入的帮扶单位类型比例

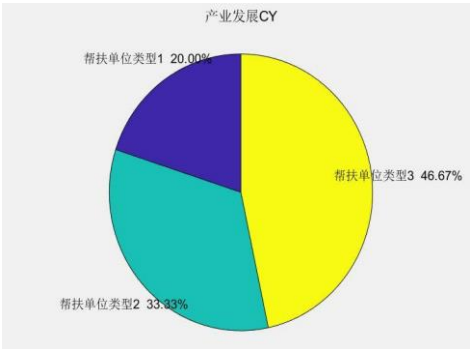

图8：产业发展上的帮扶单位类型比例

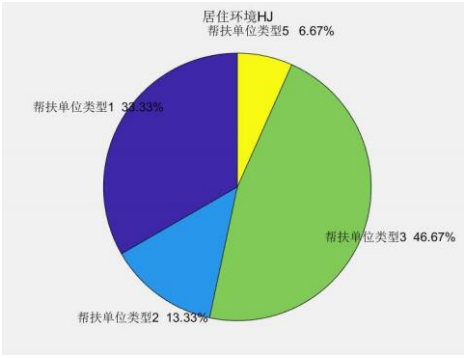

图9：居住环境的帮扶单位类型比例

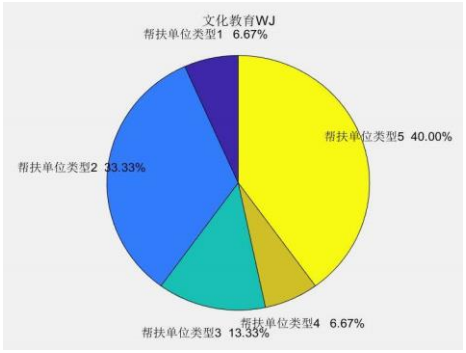

图10：文化教育的帮扶单位类型比例

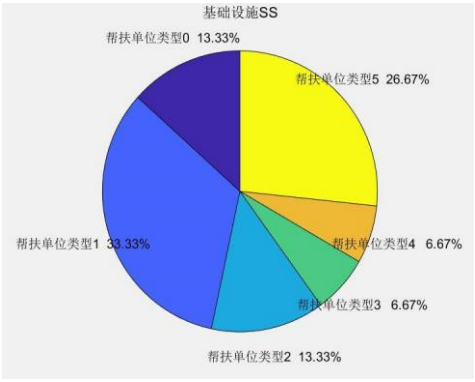

图11：基础设施的帮扶单位类型比例

6.4 问题四模型的建立与求解

6.4.1 问题四模型的建立

问题四首先要求我们给出影响“脱贫先进村庄”这一称号的因素，我们选择五项评分指标的增长率与总分数据的增长率作为影响该称号的因素。考虑到存在着2015年与2020年的评分数据或总分数据都较低的情况，这样虽然可能其数据的增长率较高，但总的脱贫效果并不好。所以为了判断最后10个村庄能否评上该称号，我们仅考虑2020年的总分数据来进行评判，总分高的即代表当前改村庄的脱贫效果好，故以2020年的总分数据作为评判能否获得“脱贫先进村庄”这一称号的标准。

我们选择2015年的六个指标数据以及帮扶单位的编号为输入，2020年的总分数据为输出，利用神经网络算法，建立预测模型，通过预测出的该10个村庄2020年的总分数据，与总分数据排名前10000名的村庄相比较，从而判断他们是否能评上该称号。由于一级称号与二级称号的比例为1：3，选择总分数据前2500名的村庄获得一级称号，总分数据后7500名的村庄获得二级称号。那么这10个村庄中总分数据位于前2500名则能评上“脱贫先进一级村庄”的称号。

●模型的准备

通过查阅文献可知[5]，BP 神经网络算法其原理是在梯度下降法，利用梯度搜索技术，以期使网络的实际输出值和期望输出值的误差均方差为最小。其优点在于泛化能力、自学习和自适应能力强，及特别适合于求解内部机制复杂的问题。

BP 神经网络的过程主要分为两个阶段，第一阶段是信号的前向传播，从输入层经过隐含层，最后到达输出层；第二阶段是反向传播，从输出层到隐层，最后到输入层，依次调节隐层到输出层的权重和偏置，输入层到隐层的权重和偏置。

首先，初始化权重，将样本模式计数器 n 和训练次数计数器设置为 1，误差 E 设置为 0，其次是输入样本并计算输出和误差，接着根据误差调制各层的权值，当网络训练后达到精度 $E_{\min}$（设为一个较小的正数），结束神经网络参数训练。神经网络基本结构图如下：

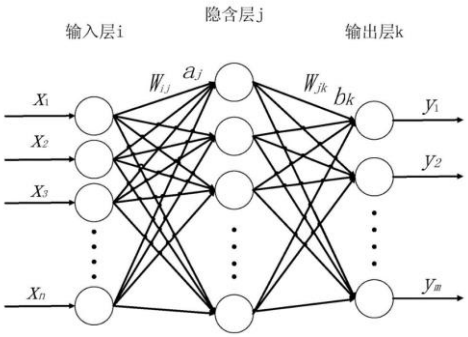

图12：神经网络基本结构图

• 模型的建立

我们选择 2015 年的六个指标数据以及帮扶单位的编号为输入，2020 年的总分数作为输出，利用神经网络算法，建立预测模型，该模型的流程结构如下图所示：

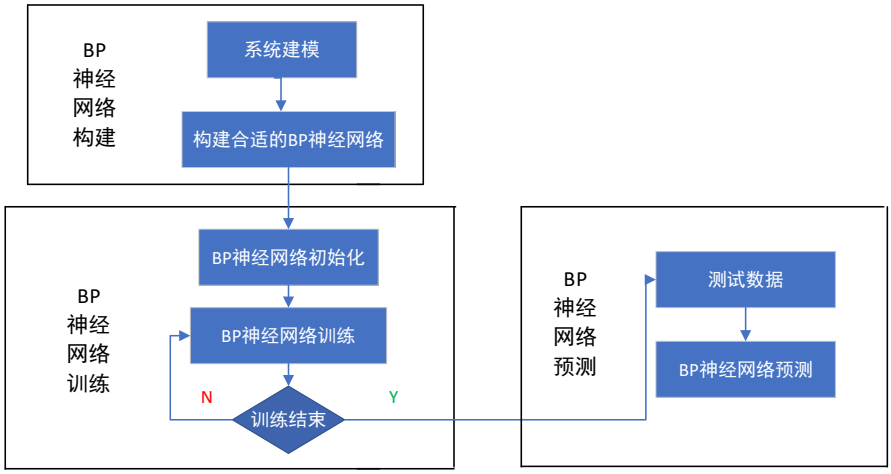

图13：预测模型流程图

在神经网络中，有如下的一些参数标识：

（1）网络的层数 n。第 l 层记为 L_l，则该神经网络中，输入层为 L_1，隐含层为 L_2, L_3，输出层为 L_4。

（2）网络权重和偏置 $(W,b)=(W^{(1)},b^{(1)},W^{(2)},b^{(2)})$，其中 $W_{ij}^{(l)}$ 表示的是第 l 层的第 j 个神经元和第 l+1层的第 i 个神经元之间的连接参数，b_i^l 标识的是第 l+1层的第 i 个神经元的偏置项。

本模型的神经网络结构如下：

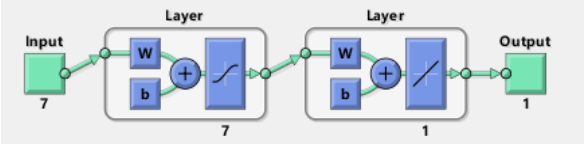

图 14：神经网络结构

隐藏层传输函数选择双曲正切 S 形函数：$a=\frac{e^{n}-e^{-n}}{e^{n}+e^{-n}}$

输出层传输函数采用线性函数：$a=n$

隐藏层神经元个数对 BP 神经网络预测精度有显著的影响，节点数太少,网络不能很好地学习,需要增加训练次数,训练的精度也受影响；节点数太多,训练时间增加,网络容易过拟合。我们参考如下公式来确定最适隐藏层神经元个数：

$$\begin{cases} l<n-1 \\ l<\sqrt{(m+n)}+a \\ l=\log_2 n \end{cases} \qquad (6\text{-}4\text{-}1)$$

式中,n为输入层节点数；l为隐含层节点数；m为输出层节点数；a为0~10之间的常数。在实际问题中,隐含层节点数的选择首先是参考公式来确定节点数的大概范围,然后用试凑法确定最佳的节点数。经过多次试验，我们选择 N=100，此时 BP 神经网络达到了较高的精度。

学习速度同样对 BP 神经网络具有重要影响作用，学习速度太小，网络学习缓慢，需要增加训练次数;学习速度太大，网络学习迅速，但是容易导致网络不收敛，影响训练的精度。我们最终决定学习速度为 0.01，训练次数为 10000。

我们选择的训练模式为贝叶斯正则化【6】，正则化参数等价于对参数引入先验分布，使得模型复杂度变小(缩小解空间),对于噪声以及 outliers 的鲁棒性增强(泛化能力)。整个最优化问题从贝叶斯观点来看是一种贝叶斯最大后验估计，其中正则化项对应后验估计中的先验信息，损失函数对应后验估计中的似然函数，两者的乘积即对应贝叶斯最大后验估计的形式。

贝叶斯框架表示如下：

$$P(x|D,\alpha,\beta,M)=\frac{P(D|x,\beta,M)P(x|,\alpha,M)}{P(D|\alpha,\beta,M)} \qquad (6\text{-}4\text{-}2)$$

其中，x包含网络所有权值和偏置值的向量，D为训练数据集，α，β为与密度函数相关的参数，M为所选取的网络结构，其他表达式表示如下：

$$\begin{cases} P(D|x,\beta,M)=\frac{1}{Z_D(\beta)}\exp(-\beta E_D) \\ P(x|\alpha,M)=\frac{1}{Z_W(\alpha)}\exp(-\alpha E_W) \\ P(D|\alpha,\beta,M)=\frac{1}{Z_F(\alpha,\beta)}\exp(-F(x)) \end{cases} \qquad (6\text{-}4\text{-}3)$$

此式分别表示训练数据D的概率密度、权值x的概率密度、训练数据D的边缘概率。

6.4.2 问题四模型的求解

该问题的整体求解思路为利用神经网络算法预测最后10个村庄2020年的总分数据，通过预测出数据，与总分数据排名前10000名的村庄相比较，判断出能评上“脱贫先进村庄”称号的村庄。

一般情况下，在对数据进行分析时，正态$Q-Q$图越接近于一条直线，R值越接近于 1，模型可信度越高。通过分析，本模型得到相关对应关系如下图所示：

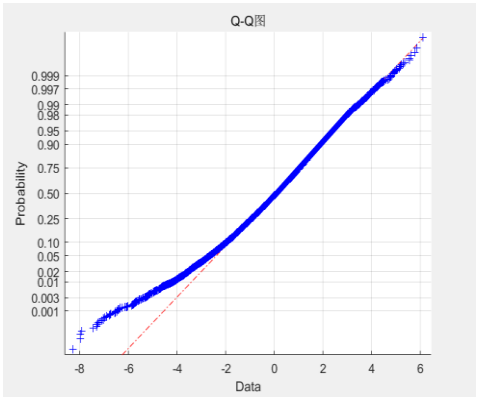

图 15：正态$Q-Q$图

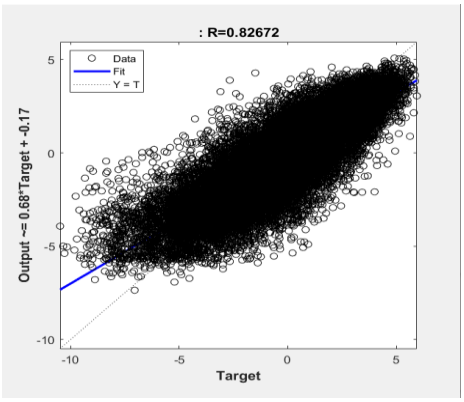

图 16：R值

正态$Q-Q$图中，数据中一串数目的每个点都是该数据的某分位点，把这些点（称为样本分位数点）和相应的理论上的分位数配对做出散点图，如果该数据服从正态分布，那么该图看上去应该像一条直线，否则就不服从正态分布。回归值 R 代表输出和目标之间的相关性，接近于 1 则代表模型拟合的较好。

由上图分析可知利用此方法得到的模型具备一定的可信度。接下来我们利用该预测模型对其他 32155 个村庄 2020 年的总分数据进行预测，与实际值相比较，预测结果如下：

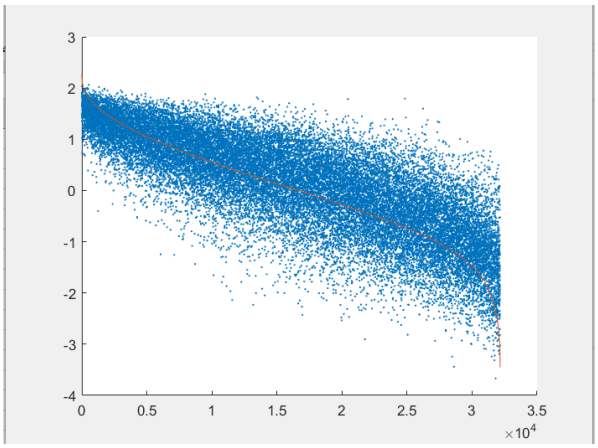

图 17：预测结果与实际结果比较图

可以看出，该预测模型的预测值与真实值基本靠近同一条直线，说明利用此方法得到的模型具备一定的准确度，再利用 matlab 工具箱对误差进行分析，其给出的误差分析相关图示如下：

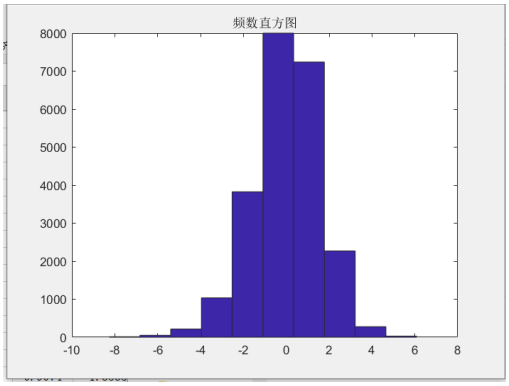

图 18：误差频数直方图

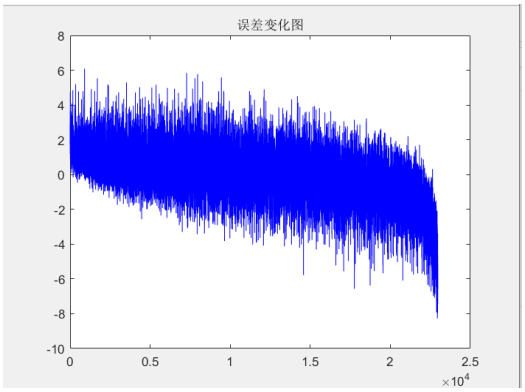

图 19：误差变化图

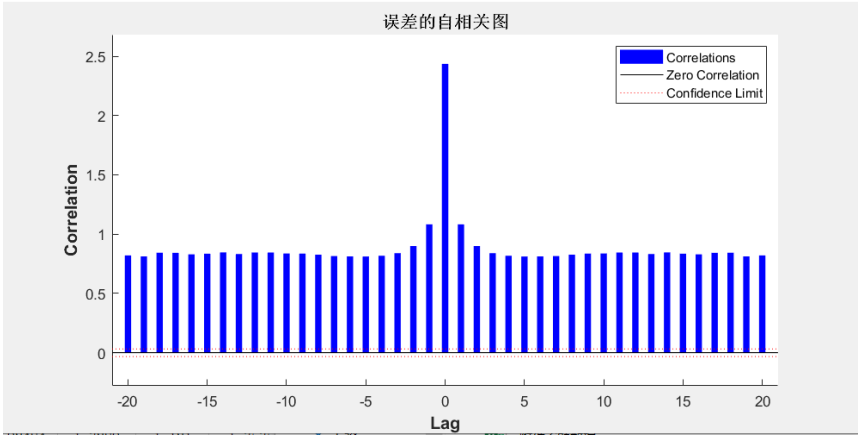

图 20：误差自相关图

通过误差分析相关图可以看出，模型的误差都比较集中，误差越集中，代表模型的效果越好，至此，我们认为该模型的可信度与准确度较高，并且有很好的预测效果。

最后，我们根据该模型得到十个村庄的预测结果如下：

表 14：十个村庄的预测结果

1	2	3	4	5
-0.9180	1.6084	0.1413	0.1993	0.2660
6	7	8	9	10
-1.9761	-1.8873	-2.3738	-0.5329	-3.5318

将预测出的总分数据与总分数据排名前 10000 名的村庄相比较，通过查阅附件可知，第 10000 名的村庄总分数据为 0.55176，故第二个村庄即编号为 25149 的村庄能评上“脱贫先进村庄”称号。由于一级称号与二级称号的比例为 1：3，通过查阅附件可知，第 2500 名村庄的总分数据为 1.3905，故该编号为 25149 的村庄能评上“脱贫先进一级村庄”。

6.5 问题五的建议信

尊敬的国家扶贫办领导：

您好！

我们数学建模组成员通过大数据对国家的脱贫帮扶绩效评价进行了相关分析。

本文基于 2015 年与 2020 年相关的评分数据以及总分数据，首先建立了典型相关模型研究了各个评价指标五年前与现在的对应关系，用数学方法，借助计算机判定了居住

环境这一指标的相关性较强，居民收入、产业发展、文化教育与基础设施的相关性较弱，但总的来说，没有相关性非常强的指标。所以五年前的评分与 2020 年对应的各项评分并没有很直接的关联。

其次，我们利用主成分分析法，建立了基于评分数据以及总分数据增长率的得分模型，得出了不同类型的帮扶单位绩效的排名以及绩效较高的帮扶单位类型，对于这些在脱贫帮扶上面有较高绩效的单位，要充分发挥出这些国有企业扶贫的优势。就单位类型而言，对于绩效较低的单位类型可以给予它们更多的建议，在政策上适当地鼓励这些企业，同时加大宣传力度与监督力度，从而更有效地解决我国贫困问题。

加上这次新冠肺炎疫情从发生至今已有一段时间，在前期的突发公共卫生事件一级响应下，使人们的生产生活不同以往，要完全恢复正常的生产生活秩序，还需要时间和耐心，其对中国经济社会发展造成负面影响已成不争事实。根据我们的研究结论，本文提出以下观点和建议：

1、要切实落实应对疫情扶持政策，把政策送到所有企业、劳动者及相关主体手中；加快落实相关政策，确保帮扶优惠到位；及时关注并顺应情势变化，调整扶持政策和措施，确保援企稳岗保发展有实效。

2、畅通农产品的流通与销售，实现稳产保供平价，既为疫情防控作出应有的贡献，又实现农户收入的稳定。

3、关注乡村旅游及休闲农业发展态势，精准帮扶[7]。

4、总结借鉴疫情防控的经验，提升贫困治理水平，对于打赢脱贫攻坚战所需要的精神和作风，需要及时加以总结和弘扬，并融入贫困治理。

同时，对于国有企业参与扶贫工作[8]我们的建议是：

1、建立扶贫常态化机制，确保扶贫企业就可以在扶贫项目中取得较大的收益。

2、寻找整合扶贫资金的途径和使用方式，这样才能发挥出扶贫资金的长期效益和整体效果。

3、做好认定扶贫龙头企业工作，实现企业与扶贫地区“共赢”的局面。

七、模型的评价与推广

7.1 模型的评价

7.1.1 模型的优点

（1）我们建立的典型相关模型，方便计算多变量之间的相关性，它的基本原理是利用两个综合变量之间的相关关系来反映两组指标之间的整体相关性，不需考虑两组变量中各个变量之间的对应关系，直接从整个考虑就可以了。

（2）我们建立的得分模型，其适用于一切因素众多，规模较大的评价问题，特别适合在社会经济系统的决策分析中使用。

（3）我们建立的预测模型，具有自适应能力，并且容错性强，能够很好的处理非线性、非局域性的大型复杂系统。

7.1.2 模型的缺点

（1）在核心指标的选区上，由于数据的匮乏，无法使用多因素聚类法对核心指标进行选取。

（2）在评价“脱贫先进村庄”时，仅考虑了 2020 年的总分数据进行评判，这样虽然说服力较强，但是综合其 2020 年的总分数据以及总分数据的增长率会使评价体系更加完善。

7.2 模型的推广

（1）我们在问题二中建立的得分模型也可以用于许多分级衡量的实际问题中，如薪资薪级，会员等级，行为属性判断等问题。

（2）我们在问题四中建立的预测模型采用了针对对象的训练模式，同理，如果改变训练模式，使其适合于其他实际问题，那么此模型也能体现出解决问题的实用性。

参考文献

[1]樊如茵.党的十八大以来扶贫工作的基本经验[J].学校党建与思想教育,2020(13):94-96.

[2]苗爱民.中国脱贫攻坚政策演进及“后 2020”时期政策调整研究[J].中共福建省委党校(福建行政学院)学报,2020(04):101-108.

[3]https://blog.csdn.net/ludan_xia/article/details/82148523?utm_medium=distribute.pc_relevant_download.none-task-blog-baidujs-1.nonecase&depth_1-utm_source=distribute.pc_relevant_download.none-task-blog-baidujs-1.nonecase

[4]司守奎，孙兆亮，数学建模算法与应用，北京：国防工业出版社，2020.

[5]孙艺铭,林雨,范佩升,张丽.基于BP神经网络的智能车电磁导航控制算法[J].科学技术创新,2020(25):103-105.

[6]Omar Albatayneh,Milhan Moomen,Ahmed Farid,Khaled Ksaibati. Complementary Modeling of Gravel Road Traffic-Generated Dust Levels Using Bayesian Regularization Feedforward Neural Networks and Binary Probit Regression[J]. International Journal of Pavement Research and Technology,2020,13(40).

[7]曹玲玲.“连片特困地区”精准扶贫多维绩效模糊综合评价——基于宿迁的调查数据[J].内蒙古科技与经济,2019(14):6-8+11.

[8]蓝磊.对国有企业参与扶贫工作的思考[J].财经界,2020(10):80-81.

附录

1、问题一源代码

```
clc;%清除命令
clear;%清空工作区
load('sj.mat');%将数据导出（2015 年-2020 年各项指标以及总分）
for i=1:1:6
x=sj(:,i);%2015 年评价指标以及总分
j=i+6;
y=sj(:,j);%2020 年评价指标以及总分
[A,B,r,U,V,stats]=canoncorr(x,y);%典型相关分析
xiangguanxing(i)=r%相关系数
jianyanbiao(i)=stats%相关系数检验表
end
```

2、问题二源代码

```
clc;%清除命令
clear;%清空工作区
load('sj.mat');
a=sj(:,1:6); %2015 年的各个指标以及总分矩阵
b=sj(:,7:12);%2020 年的各个指标以及总分矩阵
%各个指标增长率
for i=1:1:6
    for j=1:1:32155
      c(j,i)=(b(j,i)-a(j,i))/a(j,i);
    end
end
d=[c b];
r=corrcoef(d);%计算相关系数矩阵
[x,y,z]=pcacov(r)%x 的列为 r 的特征向量,y 为 r 的特征值,z 为每一个主成分的贡献率
f=repmat(sign(sum(x)),size(x,1),1);%构造与 x 同维数的元素为 1 或-1 的矩阵
x=x.*f;%修改特征向量的正负号
num=7;%选择主成分的个数
df=d*x(:,[1:num]);%计算各个主成分的得分
tf=df*z(1:num)/100;%计算综合得分
```

3、问题二画图代码

```
clc;%清除命令
clear;%清空工作区
close all;%关闭所有窗口
load('sj2.mat');
leixing=sj2(:,3);%按照综合得分排序的帮扶单位类型
danwei=sj2(:,2);%按照综合得分排序的帮扶单位
a=0;b=0;c=0;d=0;e=0;f=0;%帮扶单位类型
```

```
A=[1:1:160];%160 个帮扶单位
num=zeros(160,1);%每个帮扶单位帮助的村庄数
for i=1:1:15%综合得分前 15 名的帮扶单位类型
    if leixing(i)==0
        a=a+1;
    elseif leixing(i)==1
        b=b+1;
    elseif leixing(i)==2
        c=c+1;
    elseif leixing(i)==3
        d=d+1;
    elseif leixing(i)==4
        e=e+1;
     elseif leixing(i)==5
        f=f+1;
    end
end
Data1=[a b c d e f];
Label={'帮扶单位类型 0','帮扶单位类型 1','帮扶单位类型 2','帮扶单位类型 3','帮
扶单位类型 4','帮扶单位类型 5'};
Ratio=Data1/sum(Data1);%各个类型所对应的比例
Percentage=num2str(Ratio'*100,'%1.2f');%各个类型所对应的百分比
Percentage=[repmat(blanks(2),length(Data1),1),Percentage,repmat('%',length(
Data1),1)];
Percentage=cellstr(Percentage);%构建 cell 数组
Label=strcat(Label,Percentage');%创建标签的 cell 数组
pie(Data1,Label)%画出平面饼状图
for i=1:1:15%综合得分前 15 名的帮扶单位
    Data2=danwei(i)+1;
    for j=1:1:160
      if Data2==A(j)
          num(j)=num(j)+1;
      end
    end
end
[row,col,pinglv]=find(num);
C=[row-1 pinglv]%各帮扶单位和帮扶的村庄数
```

4、问题三源代码

```
clc;%清除命令
clear;%清空工作区
load('sj3.mat');
a=sj3(:,2:7); %2015 年的各个指标以及总分矩阵
b=sj3(:,8:13);%2020 年的各个指标以及总分矩阵
```

```
danwei=sj3(:,1);
A=[1:1:160];%160 个帮扶单位
num=zeros(160,1);%每个帮扶单位帮助的村庄数
shuju=zeros(160,6);
jieguo1=zeros(160,6);
for i=1:1:32155
    Data3=danwei(i)+1;
    for j=1:1:160
      if Data3==A(j)
          num(j)=num(j)+1;
      end
    end
end
[row,col,pinglv]=find(num);
B=[row-1 pinglv]%各帮扶单位和帮扶的村庄数
%各个指标增长率
for i=1:1:6
    for j=1:1:32155
      c(j,i)=(b(j,i)-a(j,i))/a(j,i);
    end
end
juzhen=[danwei c];
%将对应帮扶单位帮助的村庄各指标累计相加
for i=1:1:32155
    shuju(juzhen(i,1)+1,1:6)=shuju(juzhen(i,1)+1,1:6)+juzhen(i,2:7);
end
%将对应帮扶单位帮助的村庄各指标之和求平均
for j=1:1:160
    jieguo1(j,:)=shuju(j,:)./pinglv(j);
end
```

5、问题四源代码

```
clc,clear
close  all
%bp  神经网络的预测代码
%载入输出和输入数据
sj1=xlsread(&apos;C:\Users\74181\Desktop\gj1.xls&apos;);    %  把原始数据保存在纯文本文件  gj.txt 中
sj=xlsread(&apos;C:\Users\74181\Desktop\sj1.xls&apos;);    %  把原始数据保存在纯文本文件  gj.txt 中

jishu=0;
ZF2020=sj1(:,15);
sj2020(:,1:6)=sj1(:,10:15);
```

```
sj2015(:,1:6)=sj1(:,4:9);
  sj2015(:,7)=sj1(:,2);
  sjmoni(:,1:6)=sj(:,4:9);
sjmoni(:,7)=sj(:,2);
%保存数据到 matlab 的工作路径里面

%  save  p.mat;
%
%  save  t.mat;%注意 t 必须为行向量
%赋值给输出 p 和输入 t

t=ZF2020&apos;;

p=sj2015&apos;;
%数据的归一化处理，利用 mapminmax 函数，使数值归一化到[-1.1]之间
%该函数使用方法如下：[y,ps]  =mapminmax(x,ymin,ymax)，x 需归化的数据输入，
%ymin，ymax 为需归化到的范围，不填默认为归化到[-1,1]
%返回归化后的值 y，以及参数 ps，ps 在结果反归一化中，需要调用

[p1,ps]=mapminmax(p);

[t1,ts]=mapminmax(t);
%确定训练数据，测试数据,一般是随机的从样本中选取 70%的数据作为训练数据
%15%的数据作为测试数据，一般是使用函数 dividerand，其一般的使用方法如下：
%[trainInd,valInd,testInd]  =  dividerand(Q,trainRatio,valRatio,testRatio)

[trainsample.p,valsample.p,testsample.p]  =dividerand(p,0.75,0.15,0.15);

[trainsample.t,valsample.t,testsample.t]  =dividerand(t,0.75,0.15,0.15);
%建立反向传播算法的 BP 神经网络，使用 newff 函数，其一般的使用方法如下
%net  =  newff(minmax(p),[隐层的神经元的个数，输出层的神经元的个数],{隐层神
经元的传输函数，输出层的传输函数} ,&apos;反向传播的训练函数&apos;),其中 p 为
输入数据，t 为输出数据
%tf 为神经网络的传输函数，默认为&apos;tansig&apos;函数为隐层的传输函数，
%purelin 函数为输出层的传输函数
%一般在这里还有其他的传输的函数一般的如下，如果预测出来的效果不是很好，可以
调节
%TF1  =  &apos;tansig&apos;;TF2  =  &apos;logsig&apos;;
%TF1  =  &apos;logsig&apos;;TF2  =  &apos;purelin&apos;;
%TF1  =  &apos;logsig&apos;;TF2  =  &apos;logsig&apos;;
%TF1  =  &apos;purelin&apos;;TF2  =  &apos;purelin&apos;;

TF1  =  &apos;tansig&apos;;TF2  =  &apos;purelin&apos;;
net=newff(minmax(p),[7,1],{TF1  TF2},&apos;traingdm&apos;);%网络创建
```

%网络参数的设置

net.trainParam.epochs=10000;%训练次数设置

net.trainParam.goal=1e-15;%训练目标设置

net.trainParam.lr=0.01;%学习率设置,应设置为较少值，太大虽然会在开始加快收敛速度，但临近最佳点时，会产生动荡，而致使无法收敛

net.trainParam.mc=0.9;%动量因子的设置，默认为 0.9

net.trainParam.show=25;%显示的间隔次数
% 指定训练参数
% net.trainFcn = 'traingd'; % 梯度下降算法
% net.trainFcn = 'traingdm'; % 动量梯度下降算法
% net.trainFcn = 'traingda'; % 变学习率梯度下降算法
% net.trainFcn = 'traingdx'; % 变学习率动量梯度下降算法
% (大型网络的首选算法)
% net.trainFcn = 'trainrp'; % RPROP(弹性 BP)算法,内存需求最小
% 共轭梯度算法
% net.trainFcn = 'traincgf'; %Fletcher-Reeves 修正算法
% net.trainFcn = 'traincgp'; %Polak-Ribiere 修正算法,内存需求比 Fletcher-Reeves 修正算法略大
% net.trainFcn = 'traincgb'; % Powell-Beal 复位算法,内存需求比 Polak-Ribiere 修正算法略大
% (大型网络的首选算法)
%net.trainFcn = 'trainscg'; % ScaledConjugate Gradient 算法,内存需求与 Fletcher-Reeves 修正算法相同,计算量比上面三种算法都小很多
% net.trainFcn = 'trainbfg'; %Quasi-Newton Algorithms - BFGS Algorithm,计算量和内存需求均比共轭梯度算法大,但收敛比较快
% net.trainFcn = 'trainoss'; % OneStep Secant Algorithm,计算量和内存需求均比 BFGS 算法小,比共轭梯度算法略大
% (中型网络的首选算法)
%net.trainFcn = 'trainlm'; %Levenberg-Marquardt 算法,内存需求最大,收敛速度最快
% net.trainFcn = 'trainbr'; % 贝叶斯正则化算法
% 有代表性的五种算法为 :'traingdx','trainrp','trainscg','trainoss', 'trainlm'
%在这里一般是选取'trainlm'函数来训练，其算对对应的是 Levenberg-Marquardt 算法

% net.trainFcn='trainlm';
net.trainFcn='trainbr';

```
[net,tr]=train(net,trainsample.p,trainsample.t);
%计算仿真，其一般用 sim 函数

[normtrainoutput,trainPerf]=sim(net,trainsample.p,[],[],trainsample.t);%  训练的数据，根据 BP 得到的结果

[normvalidateoutput,validatePerf]=sim(net,valsample.p,[],[],valsample.t);%验证的数据，经 BP 得到的结果

[normtestoutput,testPerf]=sim(net,testsample.p,[],[],testsample.t);% 测 试 数据，经 BP 得到的结果
%将所得的结果进行反归一化，得到其拟合的数据

trainoutput=mapminmax(&apos;reverse&apos;,normtrainoutput,ts);

validateoutput=mapminmax(&apos;reverse&apos;,normvalidateoutput,ts);

testoutput=mapminmax(&apos;reverse&apos;,normtestoutput,ts);
%正常输入的数据的反归一化的处理，得到其正式值

trainvalue=mapminmax(&apos;reverse&apos;,trainsample.t,ts);%正常的验证数据

validatevalue=mapminmax(&apos;reverse&apos;,valsample.t,ts);%正常的验证的数据

testvalue=mapminmax(&apos;reverse&apos;,testsample.t,ts);%正常的测试数据
%做预测，输入要预测的数据 pnew

pnew=[313,256,239]&apos;;

pnewn=mapminmax(sj2015&apos;);

anewn=sim(net,pnewn);

anew=mapminmax(&apos;reverse&apos;,anewn,ts);
%绝对误差的计算
  pnewn2=mapminmax(sjmoni&apos;);

anewn2=sim(net,pnewn2);

anew2=mapminmax(&apos;reverse&apos;,anewn2,ts);
errors=trainvalue-trainoutput;
%plotregression 拟合图
```

```
    [MU,SIGMA,MUCI,SIGMACI]=normfit(errors,0.05)  ;
figure,plotregression(trainvalue,trainoutput)
%误差图
title(&apos;误差图&apos;)

figure,plot(1:length(errors),errors,&apos;-b&apos;)

title(&apos;误差变化图&apos;)
%误差值的正态性的检验

figure,hist(errors);%频数直方图
  title(&apos;频数直方图&apos;)
figure,normplot(errors);%Q-Q 图
    title(&apos;Q-Q 图&apos;)
[muhat,sigmahat,muci,sigmaci]=normfit(errors);%参数估计   均值,方差,均值的
0.95 置信区间,方差的 0.95 置信区间

[h1,sig,ci]=  ttest(errors,muhat);%假设检验

figure,  ploterrcorr(errors);%绘制误差的自相关图
    title(&apos;误差的自相关图&apos;)
figure,  parcorr(errors);%绘制偏相关图
  title(&apos;偏相关图&apos;)
figure,  scatter(1:32155,anew,1);
hold  on,  plot(1:32155,ZF2020);
%  figure,  scatter(1:32155,anew2,1);nFcn=&apos;trainlm&apos;;
net.trainFcn=&apos;trainbr&apos;;

[net,tr]=train(net,trainsample.p,trainsample.t);
%计算仿真，其一般用 sim 函数

[normtrainoutput,trainPerf]=sim(net,trainsample.p,[],[],trainsample.t);%  训
练的数据，根据 BP 得到的结果

[normvalidateoutput,validatePerf]=sim(net,valsample.p,[],[],valsample.t);%
验证的数据，经 BP 得到的结果

[normtestoutput,testPerf]=sim(net,testsample.p,[],[],testsample.t);% 测 试 数
据，经 BP 得到的结果
%将所得的结果进行反归一化，得到其拟合的数据

trainoutput=mapminmax(&apos;reverse&apos;,normtrainoutput,ts);

validateoutput=mapminmax(&apos;reverse&apos;,normvalidateoutput,ts);
```

```
testoutput=mapminmax(&apos;reverse&apos;,normtestoutput,ts);
%正常输入的数据的反归一化的处理，得到其正式值

trainvalue=mapminmax(&apos;reverse&apos;,trainsample.t,ts);%正常的验证数据

validatevalue=mapminmax(&apos;reverse&apos;,valsample.t,ts);%正常的验证的数据

testvalue=mapminmax(&apos;reverse&apos;,testsample.t,ts);%正常的测试数据
%做预测，输入要预测的数据 pnew

pnew=[313,256,239]&apos;;

pnewn=mapminmax(sj2015&apos;);

anewn=sim(net,pnewn);

anew=mapminmax(&apos;reverse&apos;,anewn,ts);
%绝对误差的计算
  pnewn2=mapminmax(sjmoni&apos;);

anewn2=sim(net,pnewn2);

anew2=mapminmax(&apos;reverse&apos;,anewn2,ts);
errors=trainvalue-trainoutput;
%plotregression 拟合图
    [MU,SIGMA,MUCI,SIGMACI]=normfit(errors,0.05)  ;
figure,plotregression(trainvalue,trainoutput)
%误差图
title(&apos;误差图&apos;)

figure,plot(1:length(errors),errors,&apos;-b&apos;)

title(&apos;误差变化图&apos;)
%误差值的正态性的检验

figure,hist(errors);%频数直方图
  title(&apos;频数直方图&apos;)
figure,normplot(errors);%Q-Q 图
    title(&apos;Q-Q 图&apos;)
[muhat,sigmahat,muci,sigmaci]=normfit(errors);%参数估计   均值,方差,均值的 0.95 置信区间,方差的 0.95 置信区间
```

```
[h1,sig,ci]=  ttest(errors,muhat);%假设检验

figure,  ploterrcorr(errors);%绘制误差的自相关图
    title(&apos;误差的自相关图&apos;)
figure,  parcorr(errors);%绘制偏相关图
  title(&apos;偏相关图&apos;)
figure,  scatter(1:32155,anew,1);
hold  on,  plot(1:32155,ZF2020);
%  figure,  scatter(1:32155,anew2,1);
```